

## **ABSTRACT**

A system and method are presented for the comparative analysis of textual documents. In an exemplary embodiment of the present invention the method includes accessing two or more documents, performing a linguistic analysis on each document, outputting a quantified representation of a semantic content of each document, and comparing the quantified representations using a defined metric. In exemplary embodiments of the present invention such a metric can measure relative semantic closeness or distance of two documents. In exemplary embodiments of the present invention the semantic content of a document can be expressed as a semantic vector. The format of a semantic vector is flexible, and in exemplary embodiments of the present invention it and any metric used to operate on it can be adapted and optimized to the type and/or domain of documents being analyzed and the goals of the comparison.